Making "fetch" happen: The influence of social and linguistic context on nonstandard word growth and decline

Ian Stewart and Jacob Eisenstein
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA 30318
{istewart6,jacobe}@gatech.edu

Abstract

In an online community, new words come and go: today's *haha* may be replaced by tomorrow's *lol*. Changes in online writing are usually studied as a social process, with innovations diffusing through a network of individuals in a speech community. But unlike other types of innovation, language change is shaped and constrained by the grammatical system in which it takes part. To investigate the role of social and structural factors in language change, we undertake a large-scale analysis of the frequencies of nonstandard words in Reddit. Dissemination across many linguistic contexts is a predictor of success: words that appear in more linguistic contexts grow faster and survive longer. Furthermore, social dissemination plays a less important role in explaining word growth and decline than previously hypothesized.

1 Introduction

Stop trying to make "fetch" happen! It's not going to happen! – Regina George (*Mean Girls*, 2005)

With the fast-paced and ephemeral nature of online discourse, language change in online writing is both prevalent (Androutsopoulos, 2011) and noticeable (Squires, 2010). In social media, new words emerge constantly to replace even basic expressions such as laughter: today's *haha* is tomorrow's *lol* (Tagliamonte and Denis, 2008). Why do some nonstandard words, like *lol*, succeed and spread to new contexts, while others, like *fetch*, fail to catch on? Can a word's growth be predicted from patterns of usage during its early days?

Language change can be treated like other social innovations, such as the spread of hyperlinks (Bakshy et al., 2011) or hashtags (Romero et al., 2011; Tsur and Rappoport, 2015). A key

aspect of the adoption of a new practice is its *dissemination*: is it used by many people, and in many social contexts? High dissemination enables words to achieve greater exposure among social groups (Altmann et al., 2011), and may signal that the innovation is positively evaluated.

In addition to social constraints, language change is also shaped by grammatical constraints (D'Arcy and Tagliamonte, 2015). New words and phrases rarely change the rules of the game but must instead find their place in a competitive ecosystem with finely-differentiated linguistic roles, or "niches" (MacWhinney, 1989). Some words become valid in a broad range of linguistic contexts, while others remain bound to a small number of fixed expressions. We therefore posit a structural analogue to social dissemination, which we call *linguistic dissemination*.

We compare the fates of such words to determine how linguistic and social dissemination each relate to word growth, focusing on the adoption of nonstandard words in the popular online community Reddit. The following hypotheses are evaluated:

- **H1: Nonstandard words with higher initial social dissemination are more likely to grow.** Following the intuition that words require a large social base to succeed, we hypothesize a positive correlation between social dissemination and word growth.
- **H2-weak: Nonstandard words with higher linguistic dissemination in the early phase of their history are more likely to grow.** This follows from work in corpus linguistics showing that words and grammatical patterns with a higher diversity of collocations are more likely to be adopted (Ito and Tagliamonte, 2003; Partington, 1993).

- **H2-strong: Nonstandard words with higher linguistic dissemination are more likely to grow, even after controlling for social dissemination.** This follows from the intuition that linguistic context and social context contribute differently to word growth.

To address H2, we develop a novel metric for characterizing linguistic dissemination, by comparing the observed number of n-gram contexts to the number of contexts that would be predicted based on frequency alone. Our analysis of word growth and decline includes: (1) prediction of frequency change in growth words (as in prior work); (2) causal inference of the influence of dissemination on probability of word growth; (3) binary prediction of future growth versus decline; and (4) survival analysis, to determine the factors that predict when a word's popularity begins to decline. All tests indicate that linguistic dissemination plays an important role in explaining the growth and decline of nonstandard words.

2 Related Work

Lexical change online Language changes constantly, and one of the most notable forms of change is the adoption of new words (Metcalf, 2004), sometimes referred to as lexical entrenchment (Chesley and Baayen, 2010). New nonstandard words may arise through the mutation of existing forms by processes such as truncation (e.g, *favorite* to *fave*; Grieve et al., 2016) and blending (e.g., *web+log* to *weblog* to *blog*; Cook and Stevenson, 2010). The fast pace and interconnected nature of online communication is particularly conducive to innovation, and social media provides a "birds-eye view" on the process of change (Danescu-Niculescu-Mizil et al., 2013; Kershaw et al., 2016; Tsur and Rappoport, 2015).

The most closely related work is a contemporaneous study that explored the role of weak social ties in the dissemination of linguistic innovations on Reddit, which also proposed the task of quantitatively predicting the success or failure of lexical innovations (Tredici and Fernández, 2018). One distinguishing feature of our work is the emphasis on *linguistic* (rather than social) context in explaining these successes and failures. In addition to predicting the binary distinction between success and failure, we also take on the more fine-grained task of predicting the length of time that each innovation will survive.

Social dissemination Language changes as a result of transmission across generations (Labov, 2007) as well as diffusion across individuals and social groups (Bucholtz, 1999). Such diffusion can be quantified with *social dissemination*, which Altmann et al. (2011) define as the count of social units (e.g., users) who have adopted a word, normalized by the expected count under a null model in which the word is used with equal frequency across the entire population. Altmann et al. (2011) use dissemination of words across forum users and threads to predict the words' change in frequency in Usenet, finding a positive correlation between frequency change and both kinds of social dissemination. In contrast, Garley and Hockenmaier (2012) use the same metric to predict the growth of English loanwords on German hip-hop forums, and find that social dissemination has less predictive power than expected. We seek to replicate these prior findings, and to extend them to the broader context of Reddit.

Linguistic dissemination In historical linguistics, the distribution of a new word or construction across lexical contexts can signal future growth (Partington, 1993). Furthermore, grammatical and lexical factors can explain a speaker's choice of linguistic variant (Ito and Tagliamonte, 2003; Cacoullos and Walker, 2009) and can provide more insight than social factors alone. Our study proposes a generalizable method of measuring the dissemination of a word across lexical contexts with *linguistic* dissemination and compares social and linguistic dissemination as predictors of language change.

3 Data

Our study examines the adoption of words on social media, and we focus on Reddit as a source of language change. Reddit is a social content sharing site separated into distinct sub-communities or "subreddits" that center around particular topics (Gilbert, 2013). Reddit is a socially diverse and dynamic online platform, making it an ideal environment for research on language change (Kershaw et al., 2016). Furthermore, because Reddit data is publicly available we expect that this study can be more readily replicated than a similar study on other platforms such as Facebook or Twitter, whose data is less easily obtained.

	Total	Monthly mean
Comments	1,625,271,269	45,146,424
Tokens	56,674,728,199	1,574,298,006
Subreddits	333,874	48,786
Users	14,556,010	2,302,812
Threads	102,908,726	3,079,780

Table 1: Data summary statistics.

We analyze a set of public monthly Reddit comments[1] posted between 1 June 2013 and 31 May 2016, totalling $T = 36$ months of data. This dataset has been analyzed in prior work (Hessel et al., 2016; Tan and Lee, 2015) and has been noted to have some missing data (Gaffney and Matias, 2018), although this issue should not affect our analysis. To reduce noise in the data, we filter all comments generated by known bots and spam users[2] and filter all comments created in well-known non-English subreddits.[3] The final data collected is summarized in Table 1.

We replace all references to subreddits and users (marked by the convention *r/subreddit* and *u/user*) with *r/SUB* and *u/USER* tokens, and all hyperlinks with a *URL* token. We also reduce all repeated character sequences to a maximum length of three (e.g., *loooool* to *loool*). The final vocabulary includes the top 100,000 words by frequency.[4] We replace all OOV words with UNK tokens, which comprise 3.95% of the total tokens.

3.1 Finding growth words

Our work seeks to study the growth of nonstandard words, which we identify manually instead of relying on pre-determined lists (Tredici and Fernández, 2018).To detect such words, we first compute the Spearman correlation coefficient between the time steps $\{1...T\}$ and each word w's frequency time series $f_{(1:T)}^{(w)}$ (frequency normalized and log-transformed). The Spearman correlation coefficient captures monotonic, gradual growth that characterizes the adoption of nonstan-

dard words (Grieve et al., 2016; Kershaw et al., 2016).

The first set of words is filtered by a Spearman correlation coefficient above the 85th percentile ($N = 15,017$). From this set of words, one of the authors manually identified 1,120 words in set $\mathcal{G}$ ("growth") that are neither proper nouns (*berniebot, killary, drumpf*) nor standard words (*election, voting*).[5] These words were removed because their growth may be due to exogenous influence. A "standard" word is one that can plausibly be found in a newspaper article, which follows from the common understanding of newspaper text as a more formal and standard register. Therefore, a "nonstandard" word is one that cannot plausibly be found in a newspaper article, a judgment often used by linguists to determine what counts as slang (Dumas and Lighter, 1978). In ambiguous cases, one of the authors inspected a sample of comments that included the word. We validate this process by having both authors annotate the top 200 growth candidates for standard/proper versus nonstandard (binary), obtaining inter-annotator agreement of κ=0.79.

3.2 Finding decline words

To determine what makes the growth words successful, we need a control group of "decline" words, which are briefly adopted and later abandoned. Although these words may have been successful before the time period investigated, their decline phase makes them a useful comparison for the growth words. We find such words by fitting two parametric models to the frequency series.

Piecewise linear fit We fit a two-phase piecewise linear regression on each word's frequency time series $f_{(1:T)}$, which splits the time series into $f_{(1:\hat{t})}$ and $f_{(\hat{t}+1:T)}$. The goal is to select a split point $\hat{t}$ to minimize the sum of the squared error between observed frequency f and predicted frequency $\hat{f}$:

$$\hat{f}(m_1, m_2, b, t) = \begin{cases} b + m_1 t & t \leq \hat{t} \\ b + m_1 \hat{t} + m_2(t - \hat{t}) & t > \hat{t}, \end{cases} \quad (1)$$

where b is the intercept, m_1 is the slope of the first phase, and m_2 is the slope of the second phase. Decline words $\mathcal{D}_p$ ("piecewise decline") display

---

[1]From http://files.pushshift.io/reddit/comments/ (Accessed 1 October 2016).

[2]The same list used in Tan and Lee (2015): https://chenhaot.com/data/multi-community/README.txt (Accessed 1 October 2016).

[3]We randomly sampled 100 posts from the top 500 subreddits and labelled a subreddit as non-English if fewer than 90% of its posts were identified by langid.py (Lui and Baldwin, 2012) as English.

[4]We restricted the vocabulary because of the qualitative analysis required to identify nonstandard words.

[5]Code and word lists available at: https://github.com/ianbstewart/nonstandard_word_dissemination.

Word set	Examples
$\mathcal{G}$	*idk, lmao, shitpost, tbh, tho*
$\mathcal{D}_l$	*atty, eyebleach, iifym, obeasts, trashy*
$\mathcal{D}_p$	*brojob, nparent, rekd, terpers, wot*

Table 2: Examples of nonstandard words in all word sets: growth ($\mathcal{G}$), logistic decline ($\mathcal{D}_l$) and piecewise decline ($\mathcal{D}_p$).

	Clipping	Compound	Respelling	Other	Total
$\mathcal{G}$	198	334	83	505	1,120
	(17.7%)	(29.8%)	(7.4%)	(45.1%)	
$\mathcal{D}$	53	100	108	269	530
	(10.0%)	(18.9%)	(20.4%)	(50.8%)	

Table 3: Word formation category counts in growth ($\mathcal{G}$) and decline ($\mathcal{D}$) word sets.

growth in the first phase ($m_1 > 0$), decline in the second phase ($m_2 < 0$), and a strong fit between observed and predicted data, indicated by $R^2(f, \hat{f})$ above the 85th percentile (36.1%); this filtering yields 14,995 candidates.

Logistic fit To account for smoother growth-decline trajectories, we also fit the growth curve to a logistic distribution, which is a continuous unimodal distribution with support over the non-negative reals. We identify the set of candidates $\mathcal{D}_l$ ("logistic decline") as words with a strong fit to this distribution, as indicated by R^2 above the 99th percentile (82.4%), yielding 998 candidates. The logistic word set partially overlaps with the piecewise set, because some words' frequency time series show a strong fit to both the piecewise function and the logistic distribution.

Combined set We combine the sets $\mathcal{D}_p$ and $\mathcal{D}_l$ to produce a set of decline word candidates ($N = 15,665$). Next, we filter this combined set to exclude standard words and proper nouns, yielding a total of 530 decline words in set $\mathcal{D}$. Each word is assigned a split point $\hat{t}$ based on the estimated time of switch between the growth phase and the decline phase, which is the split point $\hat{t}$ for piecewise decline words and the center of the logistic distribution $\hat{\mu}$ for the logistic decline words.

Examples of both growth and decline words are shown in Table 2. The growth words include several acronyms (*tbh*, "to be honest"; *lmao*, "laughing my ass off"), while the decline words include clippings (*atty*, "atomizer"), respellings (*rekd*, "wrecked"; *wot*, "what") and compounds (*nparent*, "narcissistic parent").

We also provide a distribution of the words across word generation categories in Table 3, including compounds and clippings in similar proportions to prior work (Kulkarni and Wang, 2018). Because the growth and decline words exhibit similar proportions of category counts, we do not ex-

pect that this will be a significant confound in differentiating growth from decline.

4 Predictors

We now outline the predictors used to measure the degree of **social** and **linguistic** dissemination in the growth and decline words.

4.1 Social dissemination

We rely on the dissemination metric proposed by Altmann et al. (2011) to measure the degree to which a word occupies a specific social niche (e.g., low dissemination implies limited niche). To compute user dissemination D^U for word w at time t, we first compute the number of individual users who used word w at time t, written $U_t^{(w)}$. We then compare this with the expectation $\tilde{U}_t^{(w)}$ under a model in which word frequency is identical across all users. The user dissemination is the log ratio,

$$\log \frac{U_t^{(w)}}{\tilde{U}_t^{(w)}} = \log U_t^{(w)} - \log \tilde{U}_t^{(w)}. \quad (2)$$

Following Altmann et al. (2011), the expected count $\tilde{U}_t^{(w)}$ is computed as,

$$\tilde{U}_t^{(w)} = \sum_{u \in \mathcal{U}_t} (1 - e^{-f_t^{(w)} m_t^{(u)}}), \quad (3)$$

where $m_t^{(u)}$ equals the total number of words contributed by user u in month t, and $\mathcal{U}_t$ is the set of all users active in month t. This corresponds to a model in which each token from a user has identical likelihood $f_t^{(w)}$ of being word w. In this way, we compute dissemination for all users (D^U), subreddits (D^S) and threads (D^T) for each month $t \in \{1...T\}$.

4.2 Linguistic dissemination

Linguistic dissemination captures the diversity of linguistic contexts in which a word appears, as measured by unique n-gram counts. We compute
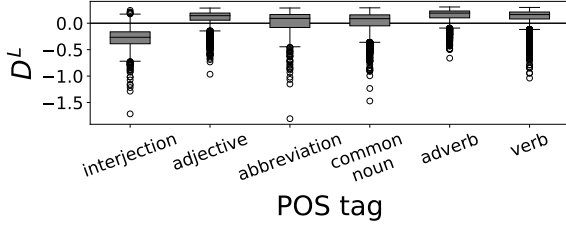
Figure 1: Distribution of mean linguistic dissemination (D^L) across part of speech groups.

the log count of unique trigram[6] contexts for all words (C^3) using all possible trigram positions: in the sentence "*that's cool af haha*", the term *af* appears in three unique trigrams, *that's cool af, cool af haha, af haha <END>*.

The unique log number of trigram contexts is strongly correlated with log word frequency ($\rho(C^3, f) = 0.904$), as implied by Heaps' law (Egghe, 2007). We therefore adjust this statistic by comparing with its expected value $\tilde{C}^3$. At each timestep t, we fit a linear regression between log-frequency and log-unique n-gram counts, and then compute the residual between the observed log count of unique trigrams and its expectation, $D^L = C_t^3 - \tilde{C}_t^3$. The residual D^L, or *linguistic dissemination*, identifies words with a higher or lower number of lexical contexts than expected.

Linguistic dissemination can separate words by grammatical category, as shown in Figure 1 where the mean D^L values are computed for words across common part-of-speech categories. Part-of-speech tags were computed over the entire corpus using a Twitter-based tagger (Gimpel et al., 2011), and each word type was assigned the most likely POS tag to provide an approximate distribution of tags over the vocabulary. Interjections have a lower median D^L than other word categories due to the tendency of interjections to occur in limited lexical contexts. Conversely, verbs have a higher median D^L due to the flexibility of verbs' arguments (e.g., subject and object may both be open-class nouns).

5 Results

The hypotheses about social and linguistic dissemination are tested under four analyses: correlation against frequency change in growth words; causal inference on probability of word growth; binary

---

[6] Pilot analysis with bigram contexts gave similar results.

prediction of word growth; and survival analysis of decline words.

5.1 Correlational analysis

To test the relative importance of the linguistic and social context on word growth, we correlate these metrics with frequency change ($\Delta_{f_t} = f_t - f_{t-k}$) across all growth words. This replicates the methodology in prior work by Altmann et al. (2011) and Garley and Hockenmaier (2012), who analyzed different internet forums. Focusing on long-term change with $k = 12$ (one year) and $k = 24$ (two years), we compute the proportion of variance in frequency change explained by the covariates using a relative importance regression (Kruskal, 1987).[7]

The results of the regression are shown in Table 4. All predictors have relative importance greater than zero, according to a bootstrap method to produce confidence intervals (Tonidandel et al., 2009). Frequency is the strongest predictor (f_{t-12}, f_{t-24}), because words with low initial frequency often show the most frequency change. In both short- and long-term prediction, linguistic dissemination (D_{t-12}^L, D_{t-24}^L) has a higher relative importance than each of the social dissemination metrics. The social dissemination metrics have less explanatory power, in comparison with the other predictors and in comparison to the prior results of Garley and Hockenmaier (2012), who found 1.5% of variance explained by D^U and 1.9% for D^T at $k = 24$. Our results were robust to the exclusion of the predictor D^L, meaning that a model with only the social dissemination metrics as predictors resulted in a similar proportion of variance explained. The weakness of social dissemination could be due to the fragmented nature of Reddit, compared to more intra-connected forums. Since users and threads are spread across many different subreddits, and users may not visit multiple subreddits, a higher social dissemination for a particular word may not lead to immediate growth.

5.2 Causal analysis

While correlation can help explain the relationship between dissemination and frequency change, it only addresses the weak version of H2: it does not

---

[7] Relative importance regression implemented in the *relaimpo* package in R: https://cran.r-project.org/package=relaimpo

	Variance explained	Lower, upper 95%
f_{t-12}	10.8%	[10.2%, 11.5%]
D^L_{t-12}	0.584%	[0.461%, 0.777%]
D^U_{t-12}	0.307%	[0.251%, 0.398%]
D^S_{t-12}	0.120%	[0.0852%, 0.191%]
D^T_{t-12}	0.246%	[0.171%, 0.379%]
f_{t-24}	21.4%	[20.4%, 22.4%]
D^L_{t-24}	1.29%	[1.05%, 1.64%]
D^U_{t-24}	0.400%	[0.346%, 0.493%]
D^S_{t-24}	0.287%	[0.201%, 0.392%]
D^T_{t-24}	0.272%	[0.226%, 0.380%]

Table 4: Percent of variance explained in frequency change, computed over all growth words $\mathcal{G}$. $N = 26,880$ for $k = 12$, $N = 13,440$ for $k = 24$.

distinguish the causal impact of linguistic and social dissemination. To test the strong version of H2, we turn to a causal analysis, in which the *outcome* is whether a nonstandard word grows or declines, the *treatment* is a single dissemination metric such as linguistic dissemination, and the *covariates* are the remaining dissemination metrics. The goal of this analysis is to test the impact of each dissemination metric, while holding the others constant.

Causal inference typically uses a binary treatment/control distinction (Angrist et al., 1996), but in this case the treatment is continuous. We therefore turn to an adapted model known as the *average dose response function* to measure the causal impact of dissemination (Imbens, 2000). To explain the procedure for estimating the average dose response, we adopt the following terminology: Z for treatment variable, X for covariates, Y for outcome.[8]

1. A linear model is fit to estimate the treatment from the covariates,

$$Z_i \mid X_i \sim \mathcal{N}(\beta^\top X_i, \sigma^2). \quad (4)$$

The output of this estimation procedure is a vector of weights $\hat{\beta}$ and a variance $\hat{\sigma}^2$.

2. The generalized propensity score (GPS) R is the likelihood of observing the treatment given the covariates, $P(Z_i \mid X_i)$. It is computed from the parameters estimated in the

---

[8] Average dose response function implemented in the *causaldrf* package in R: https://cran.r-project.org/package=causaldrf

previous step:

$$\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{(Z_i - \hat{\beta}^\top X_i)^2}{2\hat{\sigma}^2}\right). \quad (5)$$

3. A logistic model is fit to predict the outcome Y_i using the treatment Z_i and the GPS $\hat{R}_i$:

$$\hat{Y}_i = \text{Logistic}(\hat{\alpha}_0 + \hat{\alpha}_1 Z_i + \hat{\alpha}_2 \hat{R}_i). \quad (6)$$

This involves estimating the parameters $\{\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2.\}$ By incorporating the generalized propensity score $\hat{R}_i$ into this predictive model over the outcome, it is possible to isolate the causal effect of the treatment from the other covariates (Hirano and Imbens, 2004).

4. The range of treatments is divided into levels (quantiles). The *average dose response* for a given treatment level s_z is the mean estimated outcome for all instances at that treatment level,

$$\hat{\mu}(s_z) = \frac{1}{|s_z|} \sum_{z_i \in s_z} \hat{Y}_i. \quad (7)$$

The average dose response function is then plotted for all treatment levels.

Each dissemination metric is considered separately as a treatment. We consider all other dissemination metrics and frequency as covariates: e.g., for treatment variable D^L, the covariates are set to $[f, D^U, D^S, D^T]$. We bootstrap the above process 100 times with different samples to produce confidence intervals. To balance the outcome classes, we sample an equal number of growth and decline words for each bootstrap iteration.

The average dose response function curves in Figure 2 show that linguistic dissemination (D^L) produces the most dramatic increase in word growth probability. For linguistic dissemination, the lowest treatment quantile (0%-10%) yields a growth probability below 40% (significantly less than chance), as compared to the highest treatment quantile (90-100%), which yields a growth probability nearly at 70% (significantly greater than chance). This supports the strong form of H2, which states that linguistic dissemination is predictive of growth, even after controlling for the frequency and the other dissemination metrics. Subreddit dissemination also shows a mild causal effect on word growth, up to 60% in the highest treatment quantile. The other social dissemination metrics prove to have less effect on word growth.
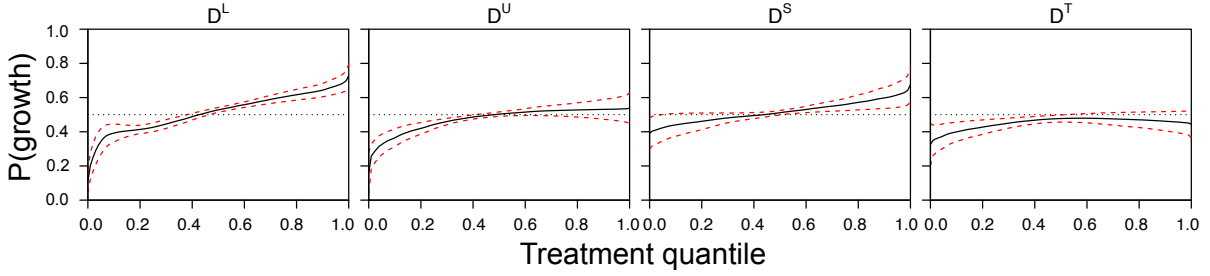
Figure 2: Average dose response function for all treatment variables, where outcome is probability of word growth. 95% confidence intervals plotted in red, chance rate of 50% marked with dotted black line.
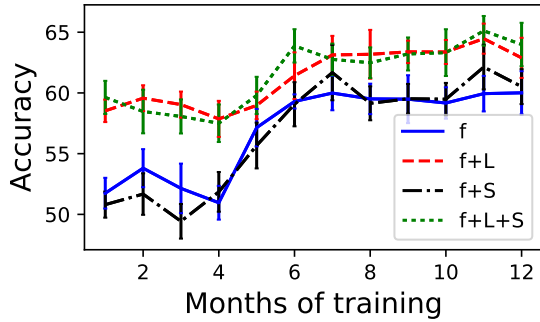

Figure 3: Prediction accuracy for different feature sets using $k = 1...12$ months of training data. f indicates frequency-only, $f+L$ frequency plus linguistic dissemination, $f + S$ frequency plus social dissemination, $f + L + S$ all features.

Figure 4: Distribution of D^L values across growth and decline words, grouped by part of speech tag. * indicates $p < 0.05$ in one-tailed t-test between growth and decline D^L values.

5.3 Predictive analysis

We now turn to prediction to determine the utility of linguistic and social dissemination: using the first k months of data, can we predict whether a word will grow or decline in popularity? This is similar to previous work in predicting the success of lexical innovations (Kooti et al., 2012), but our goal is to compare the relative predictive power of various dissemination metrics, rather than to maximize accuracy.

We use logistic regression with 10-fold cross-validation over four different feature sets: frequency-only (f), frequency plus linguistic dissemination (f+L), frequency plus social dissemination (f+S) and all features (f+L+S). Each fold is balanced for classes so that the baseline accuracy is 50%. Figure 3 shows that linguistic dissemination provides more predictive power than social dissemination: the accuracy is consistently higher for the models with linguistic dissemination than for the frequency-only and social dissemination models. The accuracies
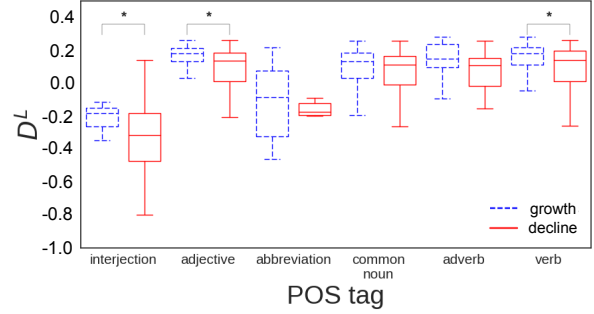
converge as the training data size increases, which suggests that frequency is a useful predictor if provided sufficient historical trajectory.

Part-of-speech robustness check Considering the uneven distribution of linguistic dissemination across part-of-speech groups (Figure 1), the prediction results may be explained by an imbalance of word categories between the growth and decline words. This issue is addressed through two robustness checks: within-group comparison and prediction.

First, we compare the distribution of linguistic dissemination values between growth and decline words, grouped by the most common POS tags (computed in § 4.2). Each decline word is matched with a growth word based on similar mean frequency in the first $k = 12$ months, and their mean linguistic dissemination values during that time period are compared, grouped within POS tag groups. The differences in Figure 4 show that across all POS tags, the growth words show a tendency toward higher linguistic dissemination with significant ($p < 0.05$) differences in the interjections and verbs.

Next, we add POS tags as additional features to the frequency-only model in the binary prediction task. The accuracy of a predictive model with access to frequency and POS features at $k = 1$ is 54.8%, which is substantially lower than the accuracy of the model with frequency and linguistic dissemination (cf. Figure 3).[9] Thus, linguistic dissemination thus contributes predictive power beyond what is contributed by part-of-speech alone.

5.4 Survival analysis

Having investigated what separates growth from decline, we now focus on the factors that precede a decline word's "death" phase (Drouin and Dury, 2009). Predicting the time until a word's decline can be framed as survival analysis (Klein and Moeschberger, 2005), in which a word is said to "survive" until the beginning of its decline phase at split point $\hat{t}$. In the Cox proportional hazards model (Cox, 1972), the hazard of death λ at each time t is modeled as a linear function of a vector of predictors,

$$\lambda_i(t) = \lambda_0(t) \exp(\beta \cdot \mathbf{x}_i), \qquad (8)$$

where $\mathbf{x}_i$ is the vector of predictors for word i, and β is the vector of coefficients. Each cell $x_{i,j}$ is set to the mean value of predictor j for word i over the training period $t = \{1...k\}$ where $k = 3$.

For words which begin to decline in popularity in our dataset, we treat the point of decline as the "death" date. The remaining words are viewed as *censored* instances: they may begin to decline in popularity at some point in the future, but this time is outside our frame of observation. We use frequency, social dissemination and linguistic dissemination as predictors in a Cox regression model.[10]

The estimated coefficients from the regression are shown in Table 5. We find a negative coefficient for linguistic dissemination ($\beta = -0.330, p < 0.001$), which mirrors the results from § 5.2: higher D^L indicates a lower hazard of word death, and therefore a higher likelihood of survival. We also find that higher subreddit dissemination has a weak but insignificant correlation with a lower likelihood of word death ($\beta = -0.156, p > 0.05$). Both of these results

---

[9]Higher k values yield similar results.

[10]Cox regression implemented in the *lifelines* package in Python: https://lifelines.readthedocs.io/en/latest/.

Predictor	β	std. error	Z	p
f	-0.207	0.0492	-4.21	***
D^L	-0.330	0.0385	-8.56	***
D^U	0.0053	0.0518	0.102	
D^S	-0.156	0.0807	-1.928	
D^T	0.0825	0.0662	1.25	

Table 5: Cox regression results for predicting word death with all predictors (f+L+S) averaged over first $k = 3$ months. *** indicates $p < 0.001$, otherwise $p > 0.05$.

lend additional support to the strong form of the hypothesis H2.

The predictive accuracy of survival analysis can be quantified by a *concordance score*. A score of 1.0 on heldout data indicates that the model perfectly predicts the order of death times; a score of 0.5 indicates that the predictions are no better than a chance ordering. We perform 10-fold cross-validation of the survival analysis model, and plot the results in Figure 5. The model with access to linguistic dissemination (f+L) consistently achieves higher concordance than the baseline frequency-only model (f), ($t = 4.29, p < 0.001$), and the model with all predictors f+L+S significantly outperforms the model with access only to frequency and social dissemination f+S ($t = 4.64, p < 0.001$). The result is reinforced by testing the goodness-of-fit for each model with model *deviance*, or difference from the null model. The f+L model has lower deviance, i.e. better fit, than the null model ($\chi^2 = 93.3, p < 0.01$), and the f+L+S does not have a significantly lower deviance than the f+L model ($\chi^2 = 4.6, p = 0.80$), suggesting that adding social dissemination does not significantly improve model fit.

6 Discussion

All four analyses support H2: linguistic dissemination was the strongest predictor of monthly frequency changes in growth words, the best differentiator of growth and decline words in causal and predictive tasks, and the most effective warning sign that a word is about to decline. Linguistic dissemination can be related to theories such as the FUDGE factors (Chesley and Baayen, 2010; Cook, 2010; Metcalf, 2004), in which a word's growth depends on frequency (F), unobtrusiveness (U), diversity of users and situations (D), generation of other forms and meanings (G), and en-

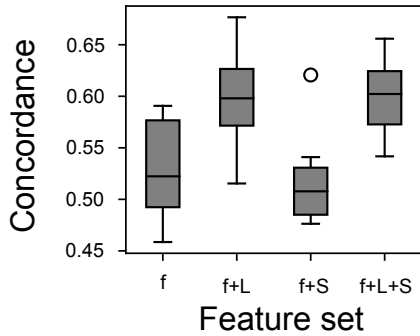Figure 5: Distribution of concordance scores (10-fold cross-validation) of the Cox regression models across feature sets.

durance (E). Linguistic dissemination provides an example of "diversity of situation."

The effectiveness of linguistic dissemination is exemplified in pairs of semantically similar growth and decline words. In the first $k = 3$ months of growth, the growth word *kinda* has a relatively high ratio of linguistic to frequency ($\frac{D^L}{f} = 0.270$) as compared with the semantically similar decline word *sorta* (0.055). This pattern holds for other pairs of semantically similar words: *fuckwit* and *fuckboy*; *lolno* and *lmao*; *yup* and *yas*. While not exhaustive, such a trend suggests that the growth words were able to reach a wider range of lexical contexts and therefore succeed where the decline words failed.

Regarding H1, we generally found a positive role for social dissemination as well, although these results were not consistent across all metrics and tests, particularly in the survival analysis. This matches the conclusion from Garley and Hockenmaier (2012), who argued that social dissemination is less predictive of word adoption than Altmann et al. (2011) originally suggested. One possible explanation is the inclusion of word categories such as proper nouns in the analysis of Altmann et al. (2011); the dissemination of such terms may rely on social dynamics more than the dissemination of nonstandard terms. The lower predictive power of thread and user dissemination is also interesting and suggests that subreddits are more socially salient in terms of exposing nonstandard words to potential adopters.

Limitations One limitation in the study was the exclusion of orthographic and morphological features such as affixation, which has been noted

as a predictor of word growth (Kershaw et al., 2016). Future work should incorporate these features as additional predictors. Our study also omitted borrowings, unlike prior work in word adoption that has focused on borrowings (Chesley and Baayen, 2010; Garley and Hockenmaier, 2012). Our early language-filtering steps eliminated most non-English words from the vocabulary, although it would have been interesting to examine loanword use in English-language posts. Finally, our study was limited by the focus on nonstandard words rather than memetic phrases (e.g., *like a boss*) which may show a similar correlation between dissemination, growth and decline (Bybee, 2006).

Future work We approximate linguistic dissemination using trigram counts, because they are easy to compute and they generalize across word categories. In future work, a more sophisticated approach might estimate linguistic dissemination with syntactic features such as appearance across different phrase heads (Kroch, 1989; Ito and Tagliamonte, 2003) or across nouns of different semantic classes (D'Arcy and Tagliamonte, 2015). Future work should also investigate more semantically-aware definitions of linguistic dissemination. The existence of semantic "neighbors" occurring in similar contexts (e.g., the influence of standard intensifier *very* on nonstandard intensifier *af*) may prevent a new word from reaching widespread popularity (Grieve, 2018).

Acknowledgments

References

Eduardo Altmann, Janet Pierrehumbert, and Adilson Motter. 2011. Niche as a determinant of word fate in online groups. *PLoS ONE*, 6(5):1–12.

Jannis Androutsopoulos. 2011. Language change and digital media: a review of conceptions and evidence. In Kristiansen Tore and Nikolas Coupland, editors, *Standard Languages and Language Standards in*

a Changing Europe, pages 145–159. Novus Press, Oslo.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. 1996. Identification of Causal Effects Using Instrumental Variables. *Source Journal of the American Statistical Association*, 91(434):444–455.

Eytan Bakshy, Jake Hofman, Winter Mason, and Duncan Watts. 2011. Everyone's an influencer: quantifying influence on Twitter. In *Proceedings of the International Conference on Web Search and Data Mining*, pages 65–74.

Mary Bucholtz. 1999. "Why be normal?": Language and identity practices in a community of nerd girls. *Language in Society*, 28:203–223.

Joan L. Bybee. 2006. From Usage to Grammar: The Mind's Response to Repetition. *Language*, 82(4):711–733.

Rena Torres Cacoullos and James Walker. 2009. The Present of the English Future: Grammatical Variation and Collocations in Discourse. *Language*, 85(2):321–354.

Paula Chesley and Harald Baayen. 2010. Predicting new words from newer words: Lexical borrowings in French. *Linguistics*, 48(6):1343–1374.

Paul Cook. 2010. *Exploiting linguistic knowledge to infer properties of neologisms*. Ph.D. thesis, University of Toronto.

Paul Cook and Suzanne Stevenson. 2010. Automatically identifying the source words of lexical blends in English. *Computational Linguistics*, 36(1):129–149.

David Cox. 1972. Regression models and life tables. *Journal of the Royal Statistical Society*, 34:187–220.

Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, and Christopher Potts. 2013. No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. *Proceedings of the International Conference on World Wide Web*, pages 307–317.

Alexandra D'Arcy and Sali Tagliamonte. 2015. Not always variable: probing the vernacular grammar. *Language Variation and Change*, 27(3):255–285.

Patrick Drouin and Pascaline Dury. 2009. When terms disappear from a specialized lexicon: A semi-automatic investigation into necrology. In *Actes de la conférence internationale "Language for Special Purposes"*.

BK Dumas and Jonathan Lighter. 1978. Is slang a word for linguists? *American Speech*, 53(1):5–17.

Leo Egghe. 2007. Untangling Herdan's law and Heaps' Law: Mathematical and informetric arguments. *Journal of the American Society for Information Science and Technology*, 58(5):702–709.

Devin Gaffney and J. Nathan Matias. 2018. Caveat emptor, computational social science: Large-scale missing data in a widely-published reddit corpus. *PLoS ONE*, 13(7).

Matt Garley and Julia Hockenmaier. 2012. Beefmoves: dissemination, diversity, and dynamics of English borrowings in a German hip hop forum. In *Proceedings of the Association of Computational Linguistics*, pages 135–139.

Eric Gilbert. 2013. Widespread Underprovision on Reddit. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pages 803–808.

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah Smith. 2011. Part-of-speech tagging for Twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 42–47.

Jack Grieve. 2018. Natural selection in the modern English Lexicon. In *International Conference on Language Evolution*, pages 153–157.

Jack Grieve, Andrea Nini, and Diansheng Guo. 2016. Analyzing lexical emergence in Modern American English online. *English Language and Linguistics*, 20(2):1–29.

Jack Hessel, Chenhao Tan, and Lillian Lee. 2016. Science, AskScience, and BadScience: On the coexistence of highly related communities. In *Proceedings of the International Conference on Social and Web Media*, pages 171–180.

Keisuke Hirano and Guido W Imbens. 2004. The propensity score with continuous treatments. In Andrew Gelman and Xiao-Li Meng, editors, *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, pages 73–84. Wiley, Chichester.

Guido W Imbens. 2000. The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.

Rika Ito and Sali Tagliamonte. 2003. Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers. *Language in Society*, 32:257–279.

Daniel Kershaw, Matthew Rowe, and Patrick Stacey. 2016. Towards Modelling Language Innovation Acceptance in Online Social Networks. *Proceedings of the ACM International Conference on Web Search and Data Mining*, pages 553–562.

John Klein and Melvin Moeschberger. 2005. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, New York.

Farshad Kooti, Winter A Mason, Krishna P Gummadi, and Meeyoung Cha. 2012. Predicting emerging social conventions in online social networks. In *Proceedings of the International Conference on Information and Knowledge Management*, pages 445–454.

Anthony S Kroch. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1(3):199–244.

William Kruskal. 1987. Relative importance by averaging over orderings. *The American Statistician*, 41(1):6–10.

Vivek Kulkarni and William Yang Wang. 2018. Simple Models for Word Formation in English Slang. In *In Proceedings of the North American Association of Computational Linguistics*, pages 1424–1434.

William Labov. 2007. Transmission and Diffusion. *Language*, 83(2):344–387.

Marco Lui and Timothy Baldwin. 2012. langid. py: An off-the-shelf language identification tool. In *Proceedings of the Association of Computational Linguistics*, pages 25–30.

Brian MacWhinney. 1989. Competition and Lexical Categorization. *Linguistic Categorization*, pages 195–242.

Allan Metcalf. 2004. *Predicting new words: The secrets of their success*. Houghton Mifflin, New York.

Alan Partington. 1993. Corpus evidence of language change: The case of the intensifier. In Mona Baker, Gill Francis, and Elena Tognini-Bonelli, editors, *Text and Technology: In Honour of John Sinclair*, pages 177–192. John Benjamins Publishing, Philadelphia.

Daniel Romero, Brendan Meeder, and Jon Kleinberg. 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter. In *Proceedings of the International Conference on World Wide Web*, pages 695–704. ACM.

Lauren Squires. 2010. Enregistering internet language. *Language in Society*, 39:457–492.

Sali Tagliamonte and Derek Denis. 2008. Linguistic ruin? LOL! Instant messaging and teen language. *American Speech*, 83(1):3–34.

Chenhao Tan and Lillian Lee. 2015. All who wander: On the prevalence and characteristics of multi-community engagement. In *Proceedings of the International Conference on World Wide Web*, pages 1056–1066.

Scott Tonidandel, James LeBreton, and Jeff Johnson. 2009. Determining the statistical significance of relative weights. *Psychological methods*, 14(4):387.

Marco Del Tredici and Raquel Fernández. 2018. The Road to Success: Assessing the Fate of Linguistic Innovations in Online Communities. In *Proceedings of the International Conference on Computational Linguistics*, pages 1591–1603.

Oren Tsur and Ari Rappoport. 2015. Don't Let Me Be #Misunderstood: Linguistically Motivated Algorithm for Predicting the Popularity of Textual Memes. In *Proceedings of the International Conference on Social and Web Media*, pages 426–435.